A 130nm Generation High Density Etox™ Flash Memory Technology

Stephen N. Keeney

Intel Corporation, Santa Clara, California, USA

Outline of Presentation

- Introduction
- Scaling Challenges
- Flash cell scaling
- Transistor Optimization
- MLC capability and embedded logic
- Results
- Conclusions
- Acknowledgements

Introduction

- A 130nm generation flash memory technology will be described.
- This technology has been optimized for a small cell size, low power, high performance and multi-level-cell operation
- The technology is designed to be compatible with a logic process

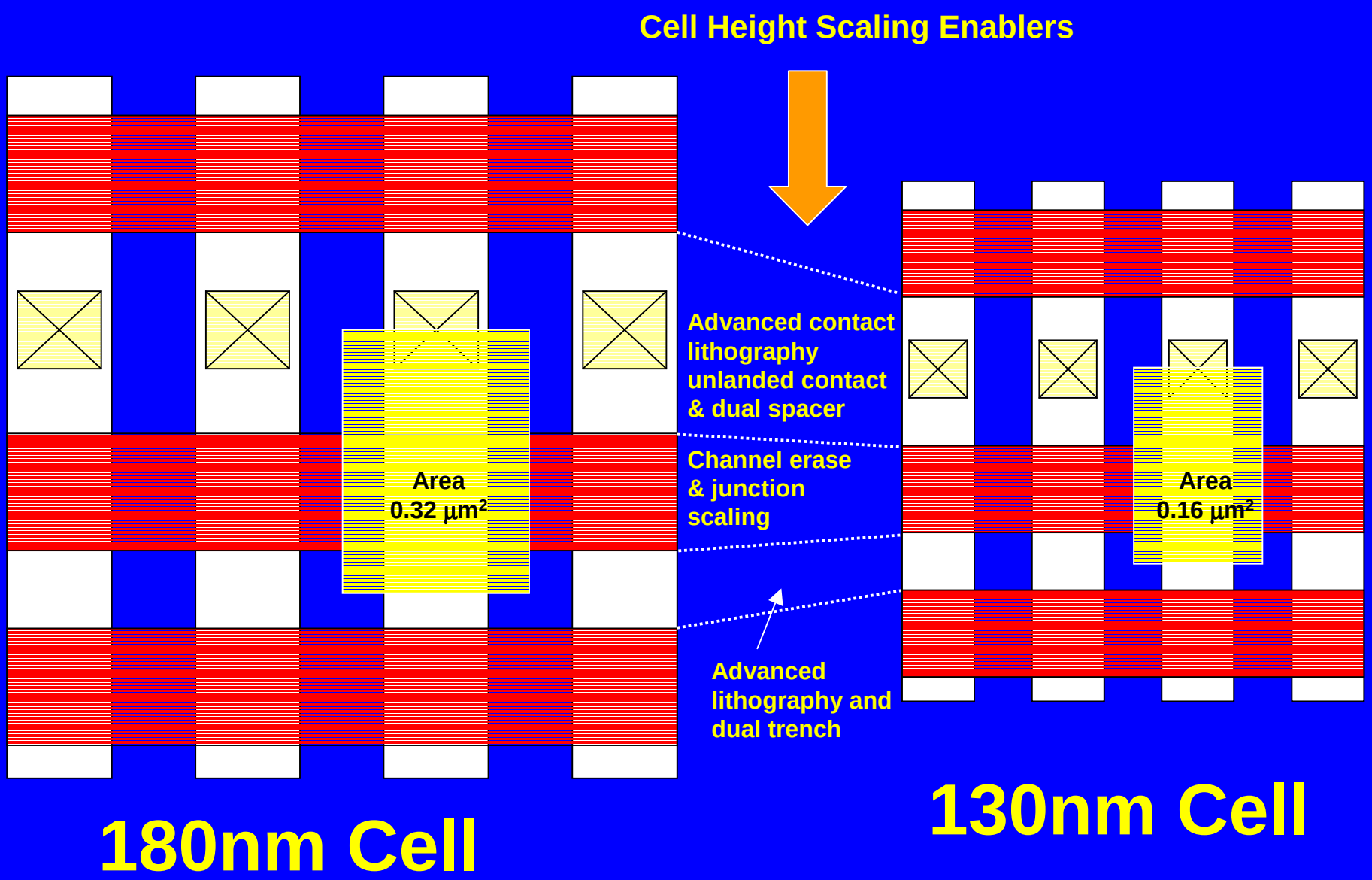
Outline of Presentation

- Introduction
- **Scaling Challenges**
- Flash cell scaling
- Transistor Optimization
- MLC capability and embedded logic
- Results
- Conclusions
- Acknowledgements

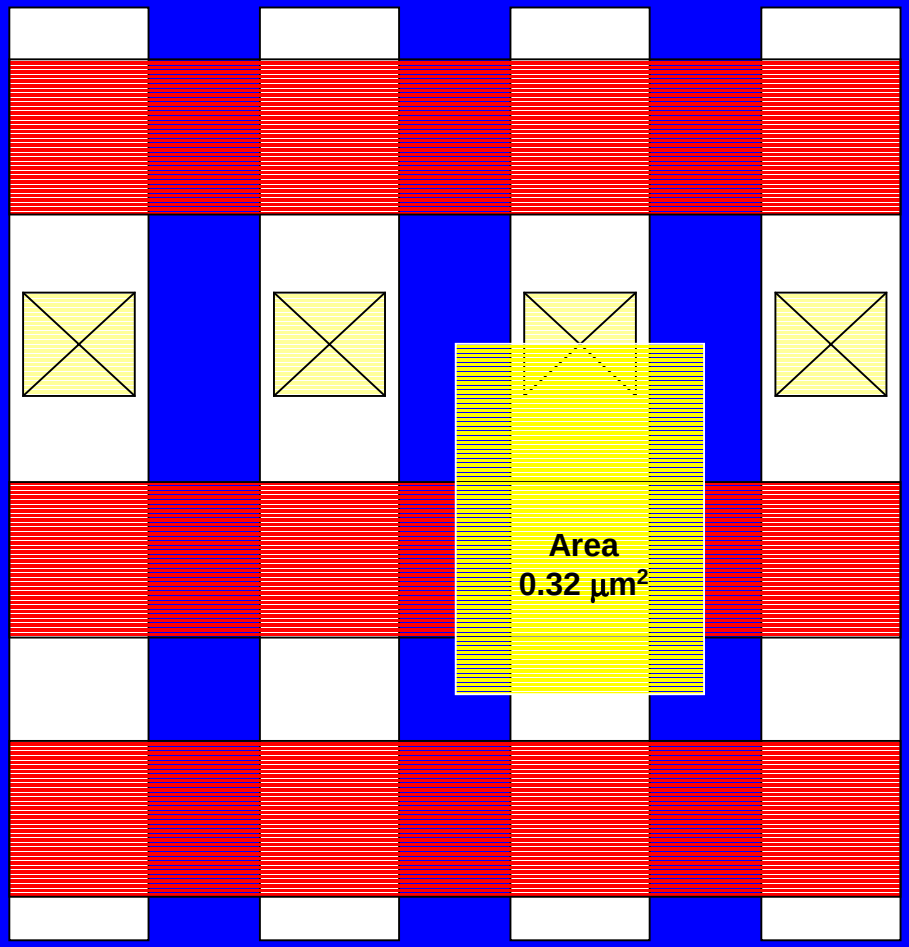
Scaling Challenges

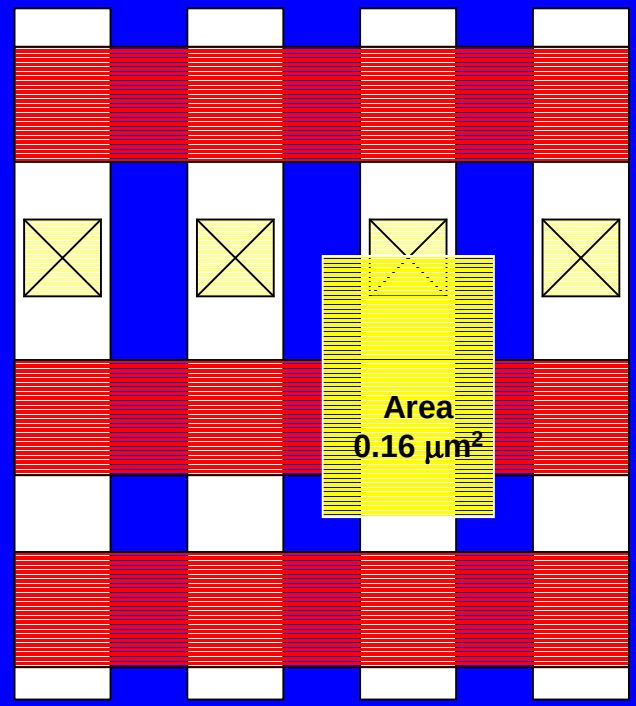
- For the flash cell the key challenges are:
  - Height scaling determined by:
    - Source space scaling
    - Gate length scaling
    - Drain space scaling
  - Width scaling determined by:
    - Electrical to physical loss in the width of the device
    - Isolation of neighboring cells
- For the periphery CMOS device the challenges are:
  - Isolation scaling, junction scaling and channel length scaling

Flash Cell Height Scaling



Flash Cell Width Scaling





Advanced lithography and scaled oxidations

Advanced lithography unlanded contact, self aligned poly and dual trench

180nm Cell

130nm Cell

Outline of Presentation

- Introduction
- Scaling Challenges
- **Flash cell scaling**
- Transistor Optimization
- MLC capability and embedded logic
- Results
- Conclusions
- Acknowledgements

Flash Cell Scaling

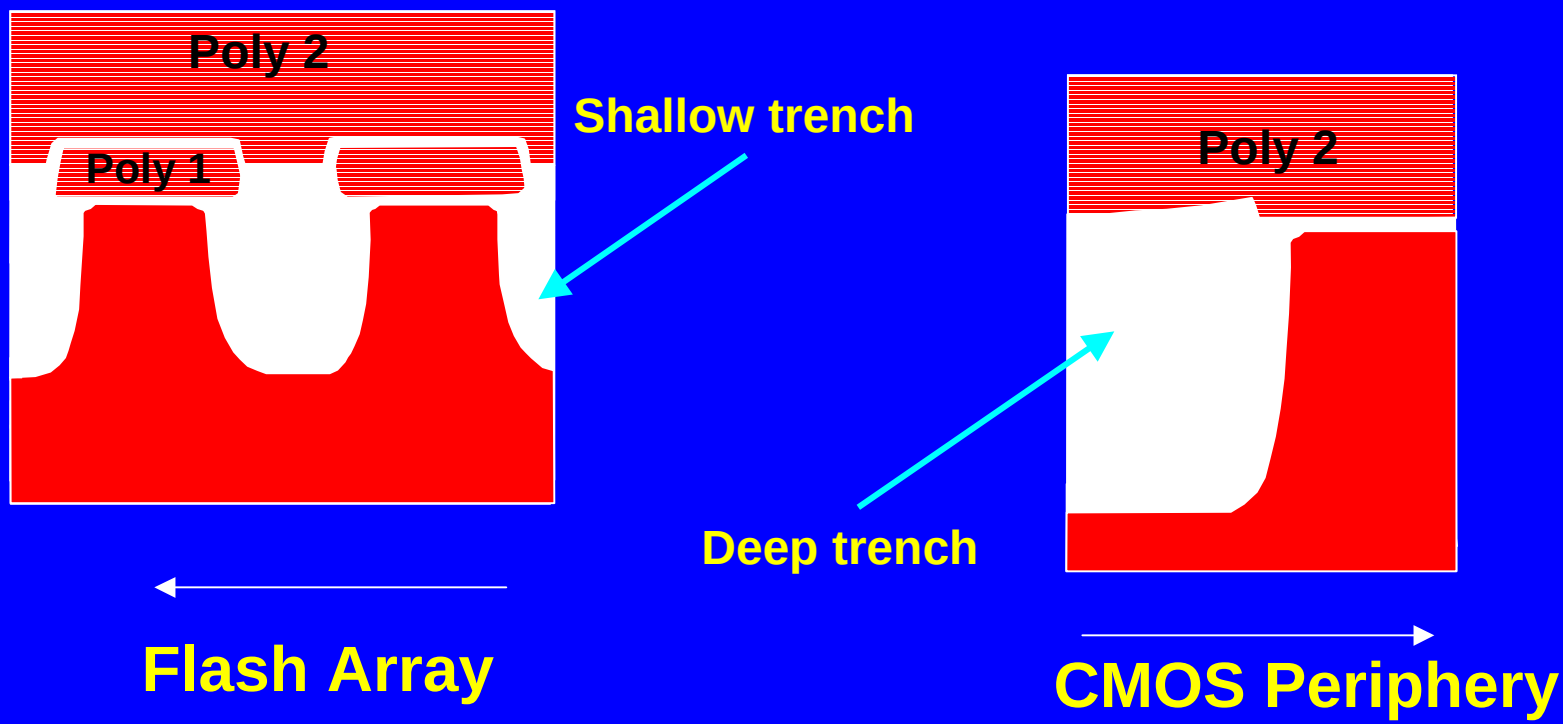
- Flash cell scaling was enabled by three key architecture components:
  - A dual trench scheme
  - A dual spacer scheme
  - A channel erase scheme

Flash Cell Scaling

- **Flash cell scaling was enabled by three key architecture components:**
  - **A dual trench scheme**
  - **A dual spacer scheme**
  - **A channel erase scheme**

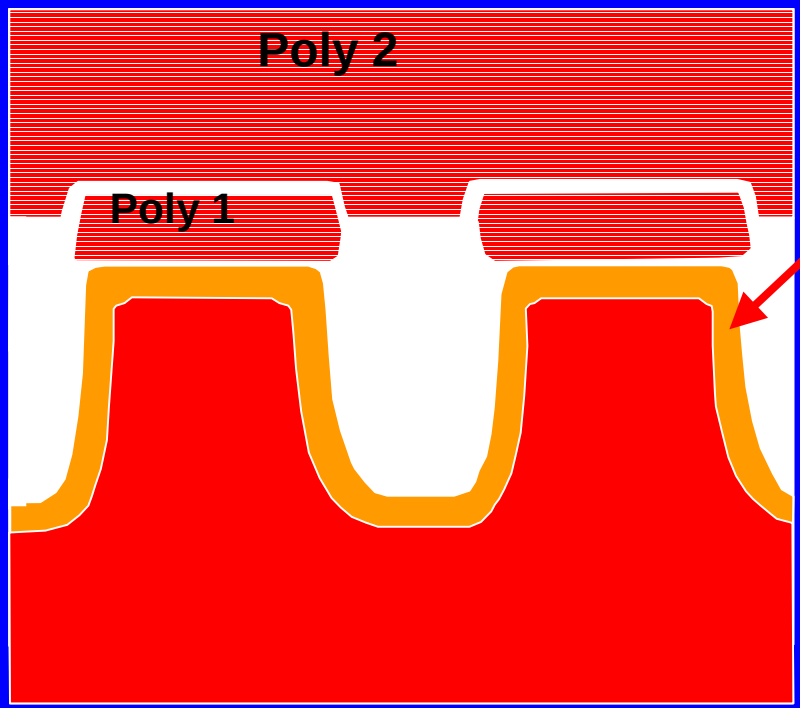
Dual Trench Scheme

- Shallow trench in the array enables cell width scaling as well as source rail width reduction
- Deep periphery supports CMOS isolation for low leakage and high voltage capability



Source Rail Resistance Scaling

- The source rail (buried diffusion) resistance is maintained equivalent to the 180nm technology by scaling the trench depth

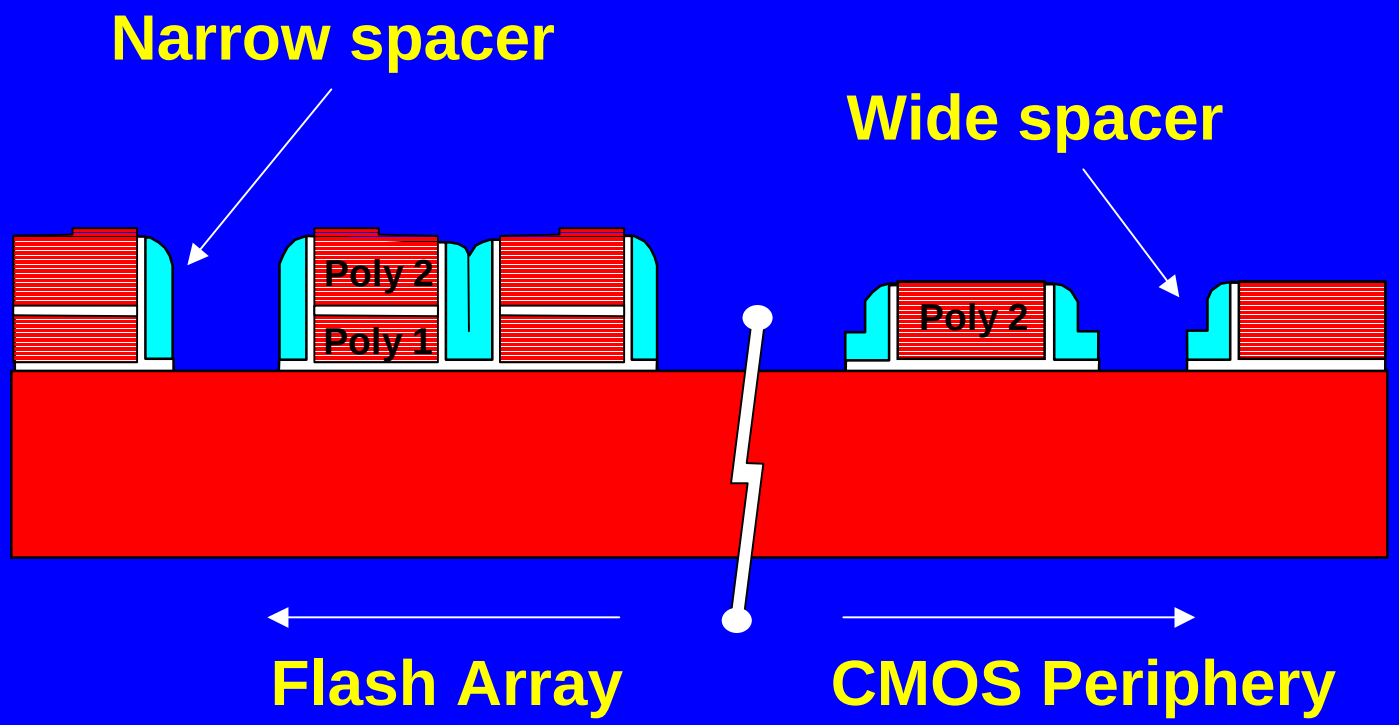


Flash Cell Scaling

- **Flash cell scaling was enabled by three key architecture components:**
  - **A dual trench scheme**
  - **A dual spacer scheme**
  - **A channel erase scheme**

Dual Spacer Scheme

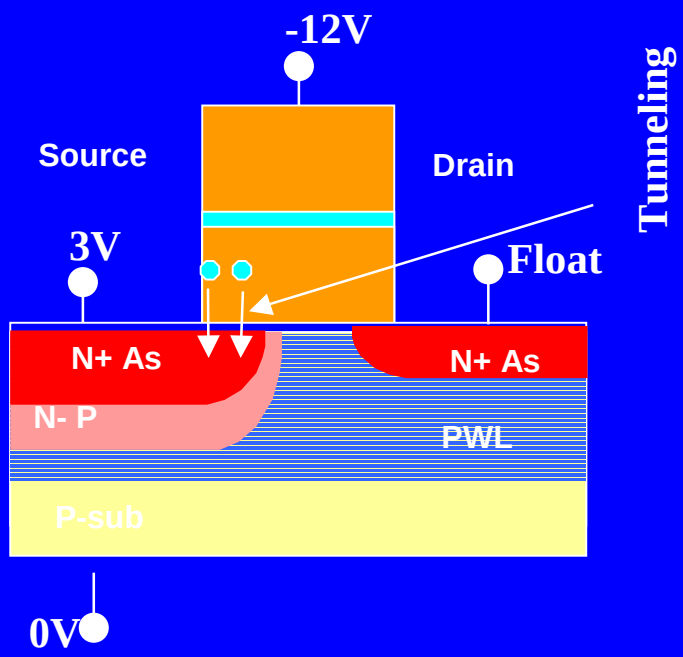
- Wide spacer required for high voltage devices in the periphery
- Narrow spacer is used elsewhere to enable poly space scaling without introducing a gap fill concern

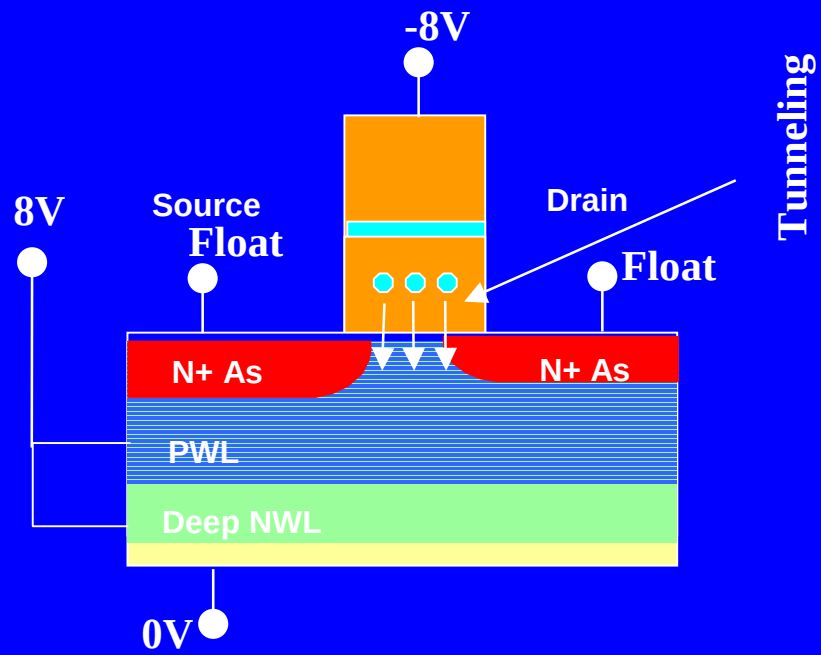


Flash Cell Scaling

- **Flash cell scaling was enabled by three key architecture components:**
  - **A dual trench scheme**
  - **A dual spacer scheme**
  - **A channel erase scheme**

Channel Erase





Source Erase Cell (180nm):

- Tunneling takes place over the Source/ Gate overlap region.
- Good Source Gate overlap required to ensure large enough Tunneling area.
- Source junction grading required to minimize the Band-to-Band Tunneling current from source.

Channel Erase Cell (130nm):

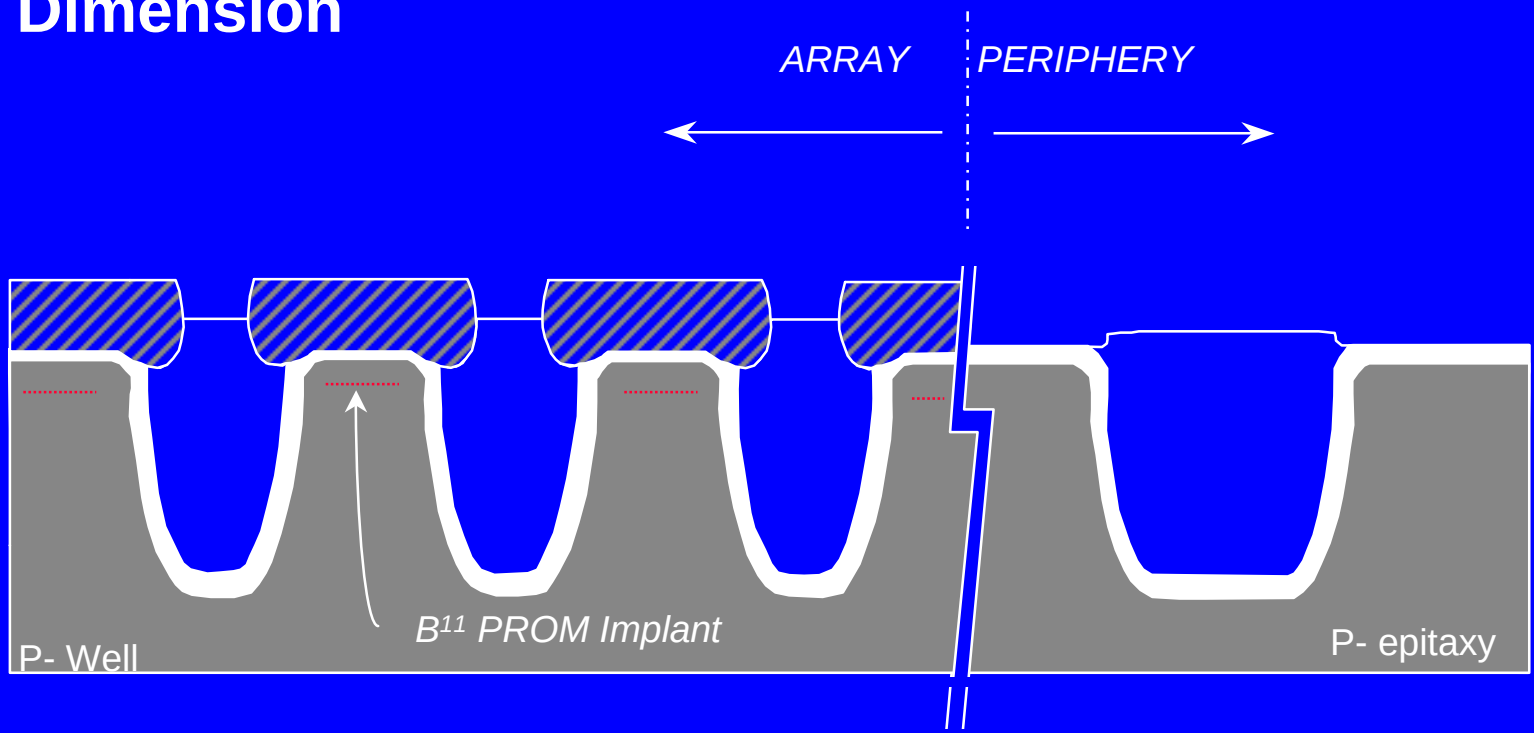
- Tunneling takes place over the entire channel region.
- Source Gate overlap can be scaled without impacting Tunneling area.
- No Band-to-Band Tunneling current from source.
- Array P-Well isolated from substrate using the Triple Well.

Architecture Scaling Features From The 180nm Generation

- This technology maintains the architecture features from the 180nm technology which are also key for scaling at the 130nm generation. These include:
  - Self aligned poly 1 (SAP)
  - Unlanded contact (ULC)
  - Cobalt Salicide
  - HDP oxides
  - Complimentary CMOS gates

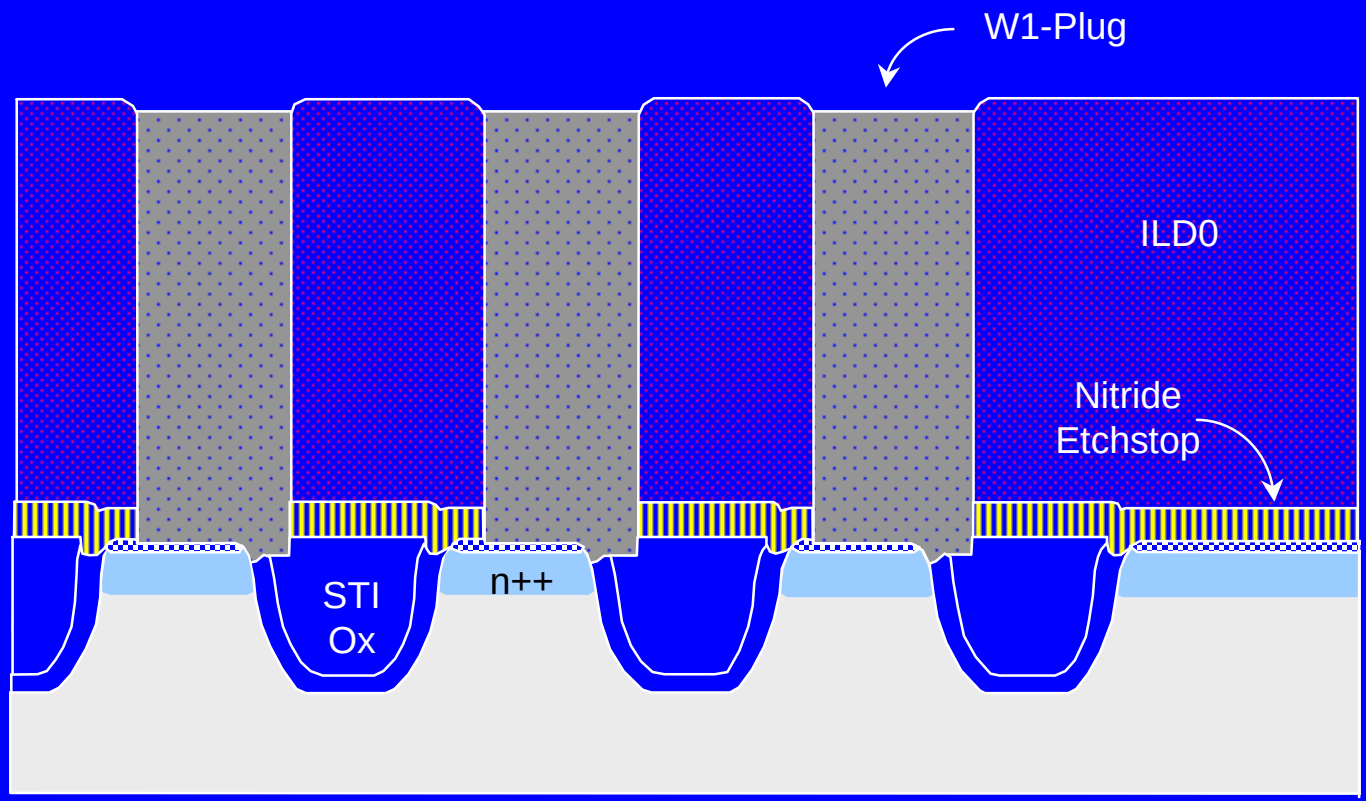
Self-Aligned Floating Gate From The 180nm Technology

- Floating Gates Self-aligned to Isolation Using CMP
- Eliminates Critical Litho Step and F.G. Space and Registration From Cell Width Minimum Dimension



Unlanded Contacts From The 180nm Technology

- Allows Contact to Mis-align and Partially Land on Trench Isolation.
- Employs Nitride Etch Stop Layer and Nitride to Oxide Etch Selectivity of >20:1



Outline of Presentation

- Introduction
- Scaling Challenges
- Flash cell scaling
- **Transistor Optimization**
- MLC capability and embedded logic
- Results
- Conclusions
- Acknowledgements

Periphery Transistor Scaling

- Periphery scaling is achieved by the following:
  - reducing the maximum voltages that need to be supported by adopting channel erase
    - Allows both junction scaling and gate oxide scaling
  - Moving to more advanced lithography for improved resolution and registration
  - More advanced etch capability
  - Traditional junction scaling

Outline of Presentation

- Introduction
- Scaling Challenges
- Flash cell scaling
- Transistor Optimization
- **MLC capability and embedded logic**
- Results
- Conclusions
- Acknowledgements

MLC Capability For Low Cost

- The technology was designed to support multi-level-cell products as well as the standard 1 bit per cell.
- For a 2 bit per cell implementation the effective cell size is 0.08um^2
- This will provide the capability for very high density low cost memory due to the small die size.

Logic Compatibility

- The technology is compatible with a logic process and is therefore suited for flash+logic embedded applications since it provides all the features for an advanced transistor including:
  - Deep periphery trench for low leakage
  - Complimentary gate CMOS
  - Cobalt Salicide
  - Gate patterning capability to 100nm
  - Thin (multiple) gates supported
  - Shallow junctions

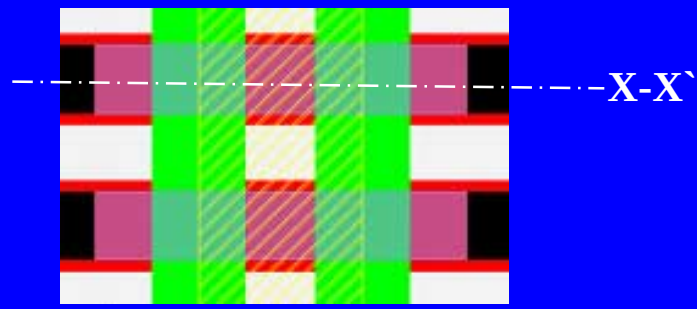
Outline of Presentation

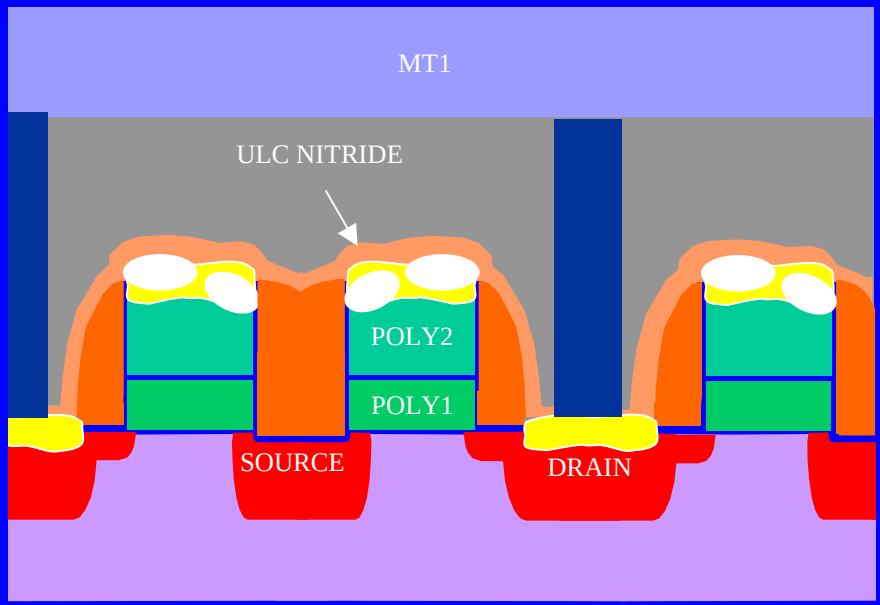
- Introduction
- Scaling Challenges
- Flash cell scaling
- Transistor Optimization
- MLC capability and embedded logic
- **Results**
- Conclusions
- Acknowledgements

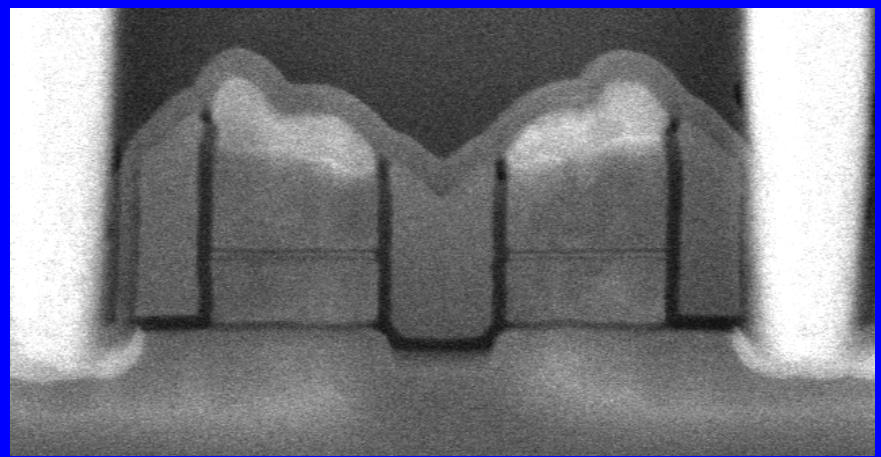
Flash Array Cross-Sections

- Wordline direction
- Bitline direction

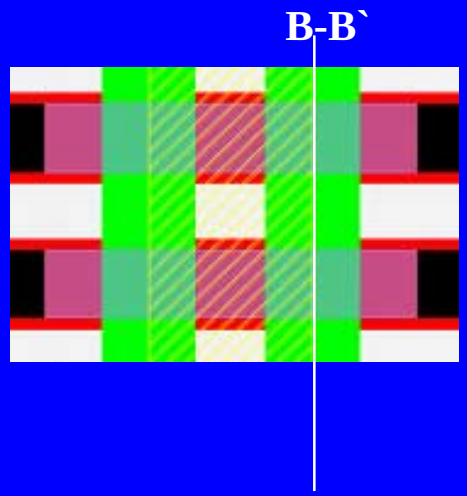
Array Cross-Section In The Bitline Direction

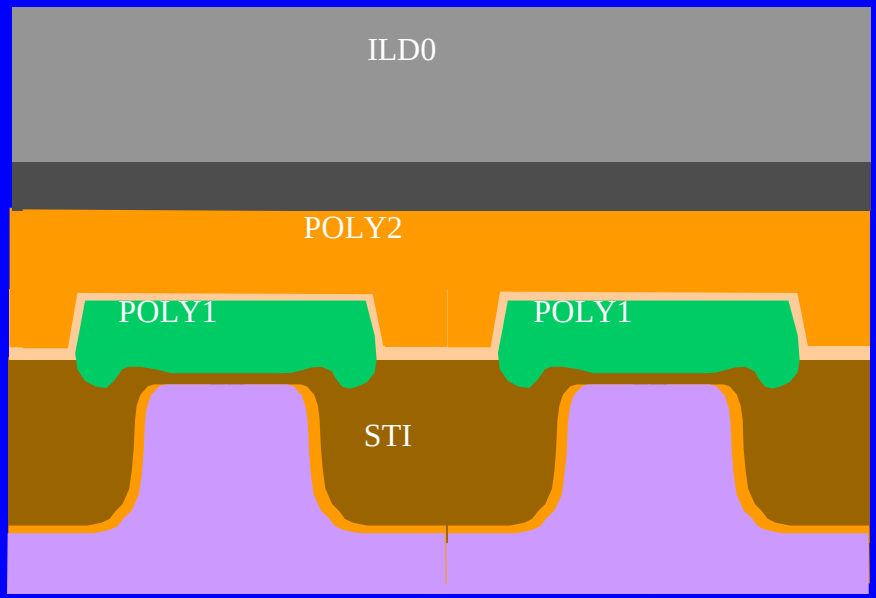


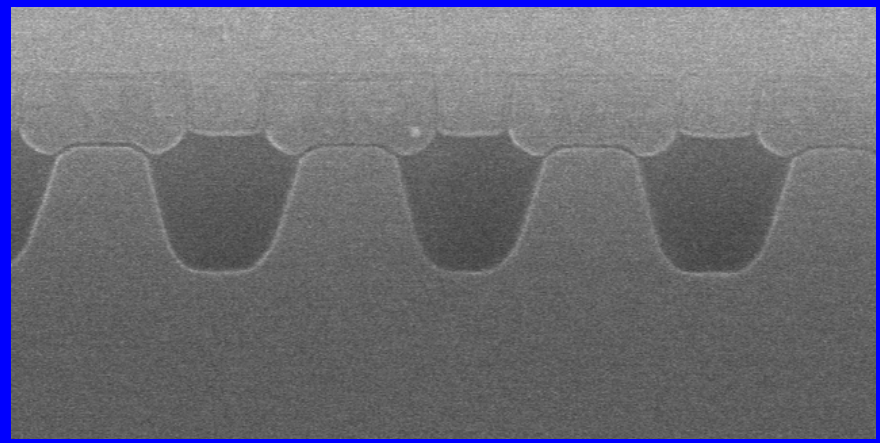




Array Cross-Section Along The Wordline





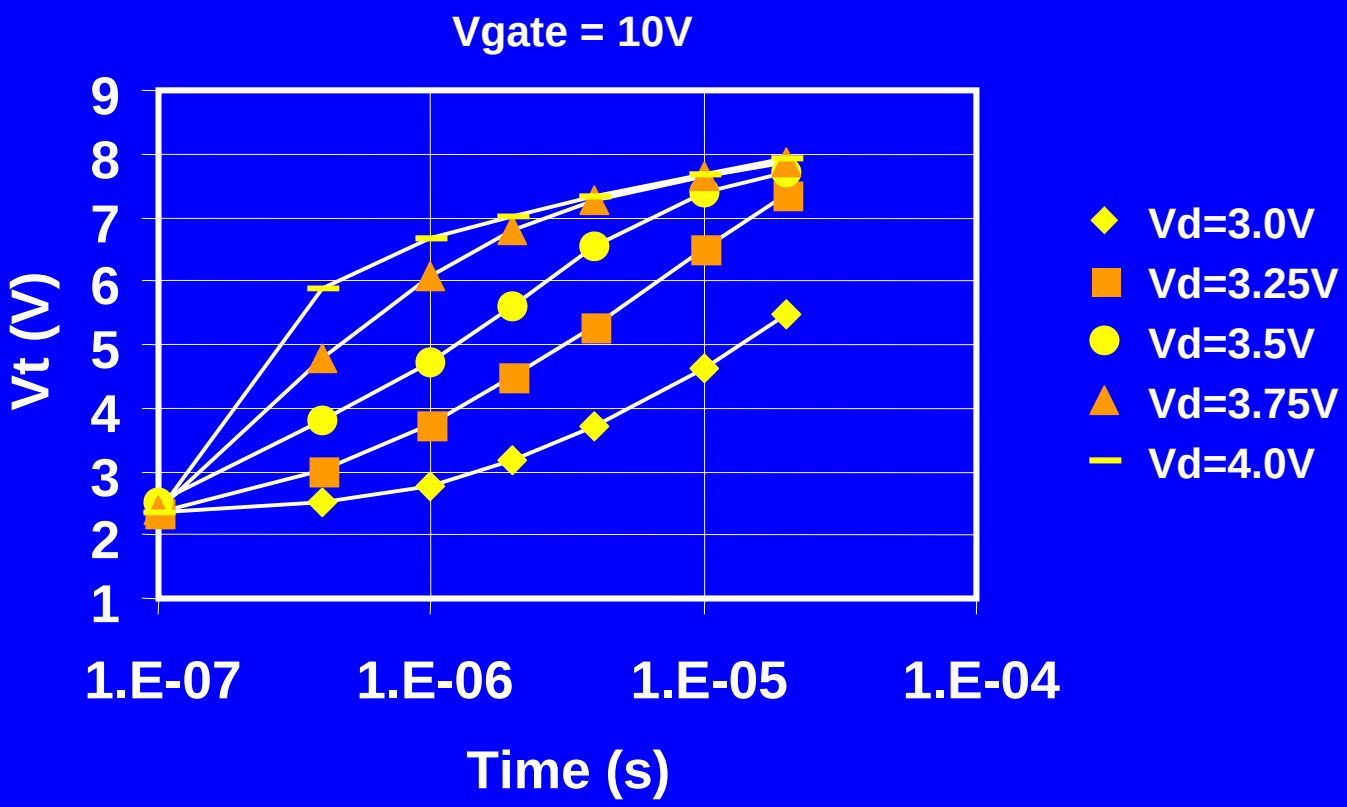


Device Performance Data

- Flash cell programming characteristics
- Flash cell erase characteristics
- Periphery transistor sub-threshold characteristics

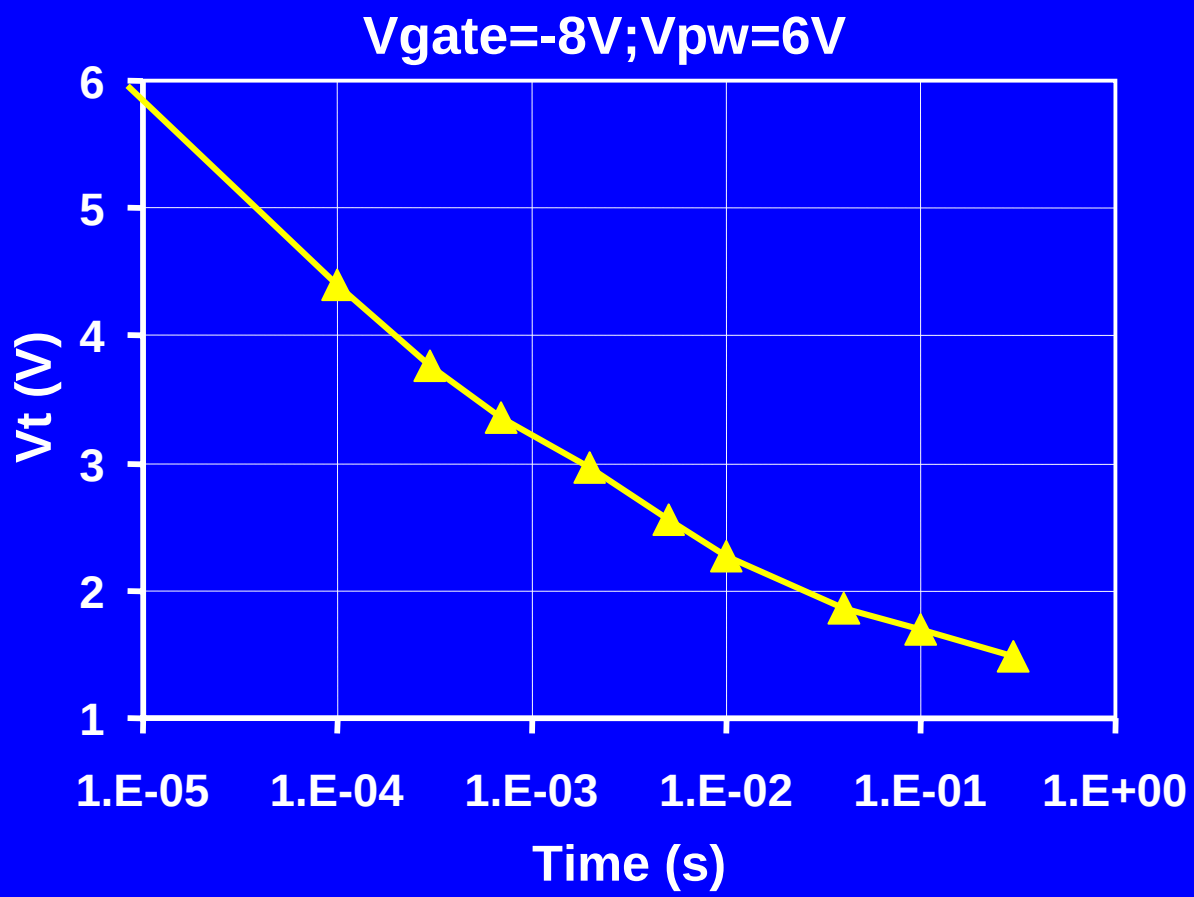
Flash Cell Programming Performance

- Good programming performance achieved at low drain voltage



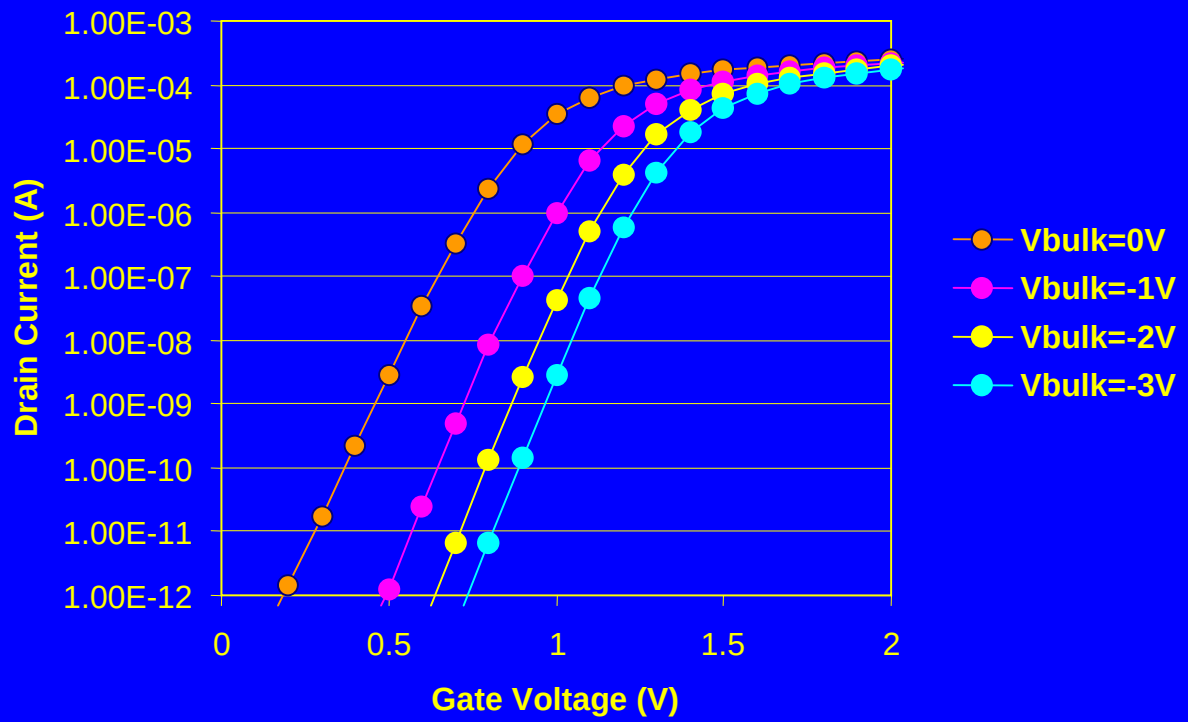
Flash Cell Erase Performance

- A typical cell will erase in ~100ms under nominal voltage conditions



Periphery Transistor Characteristics

- **Sub-threshold characteristics at minimum channel length show no kink with back bias**



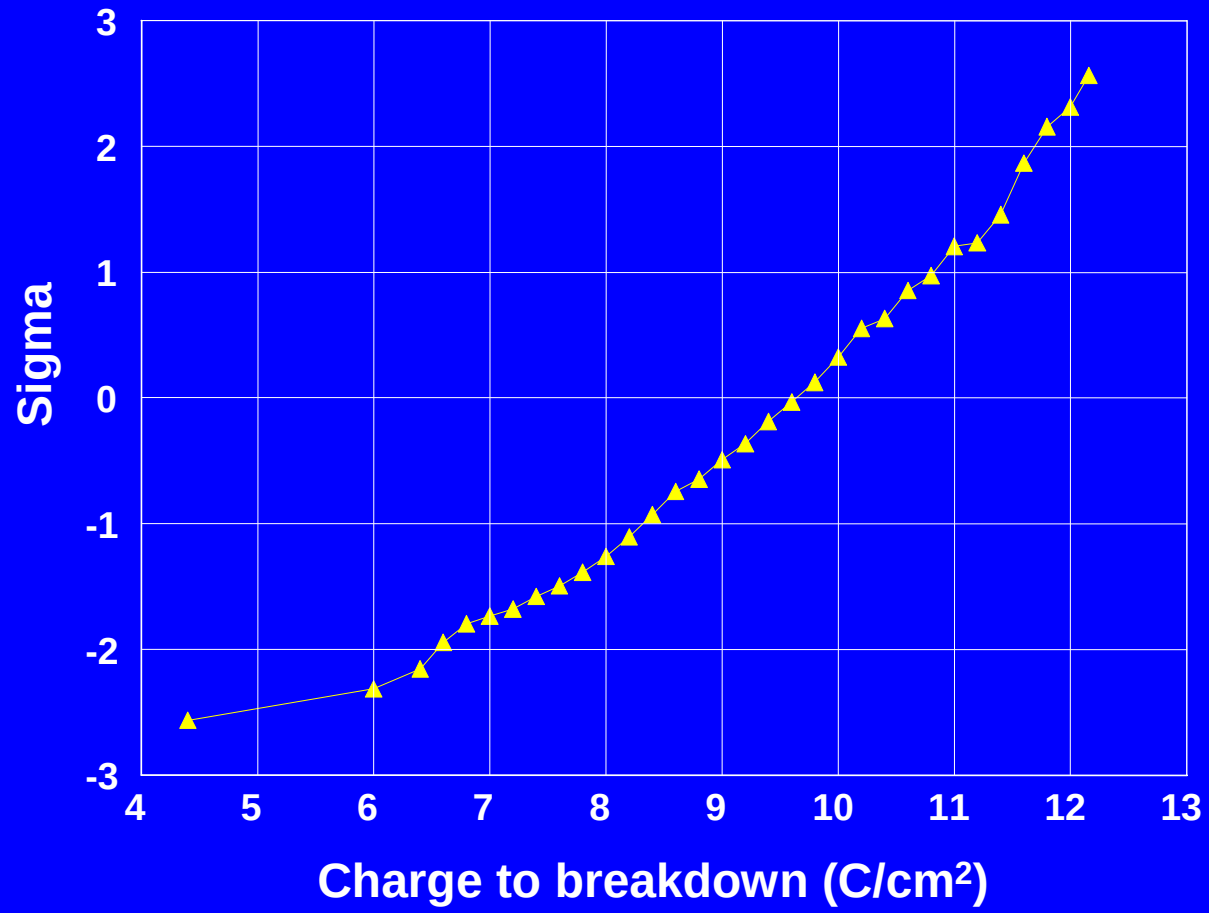
Drain Current versus Gate Bias @ Vd=50mV and Vb of 0V,-1V,-2V and -3V

Device Reliability Data

- Tunnel oxide Qbd (charge to breakdown)
- Periphery Oxide (charge to breakdown)
- Flash cell post cycling bake retention
- Flash cell endurance characteristics

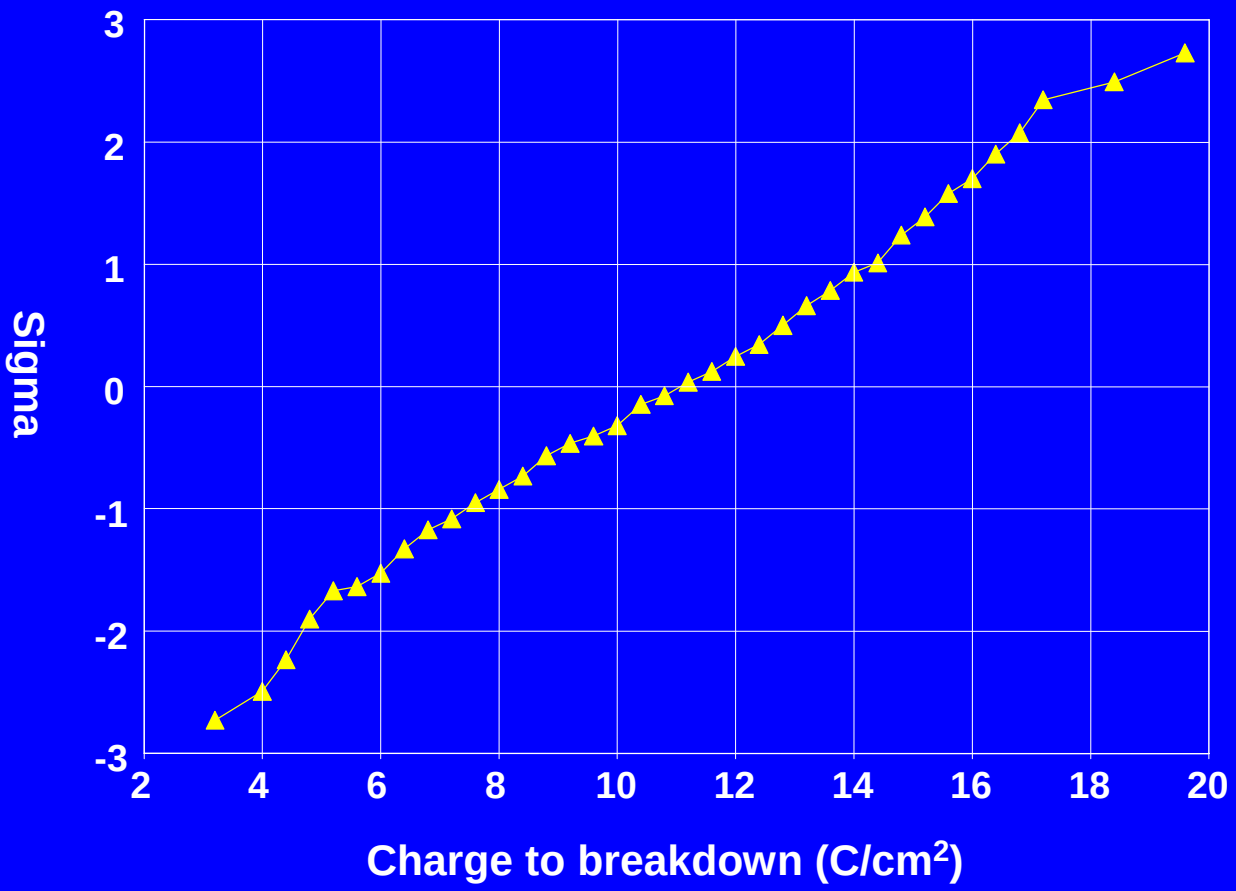
Tunnel Oxide Quality

- Good oxide quality indicating good trench corner rounding in the array



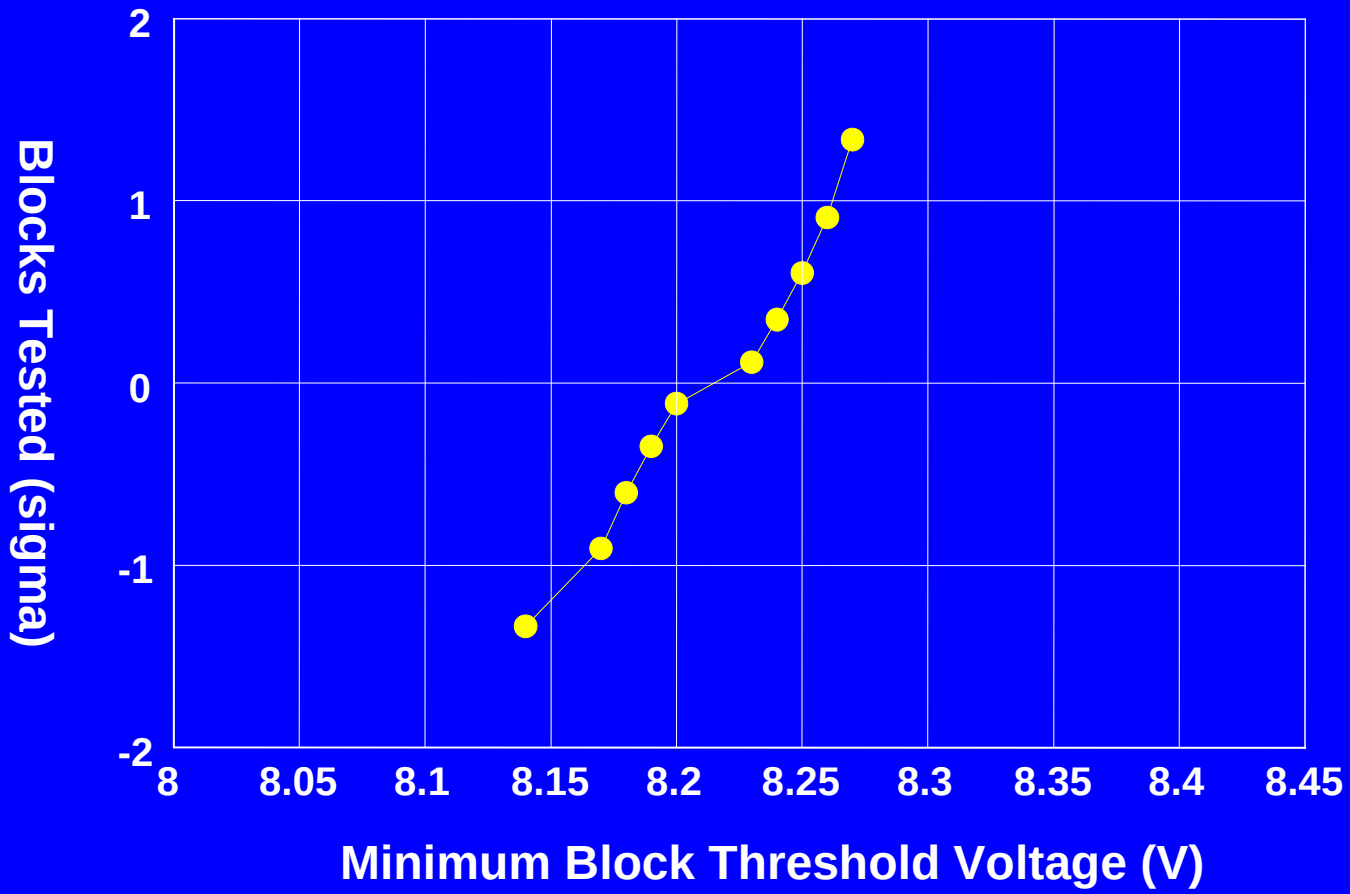
Periphery Transistor Oxide Quality

- Good oxide quality showing no trench corner impact in the periphery



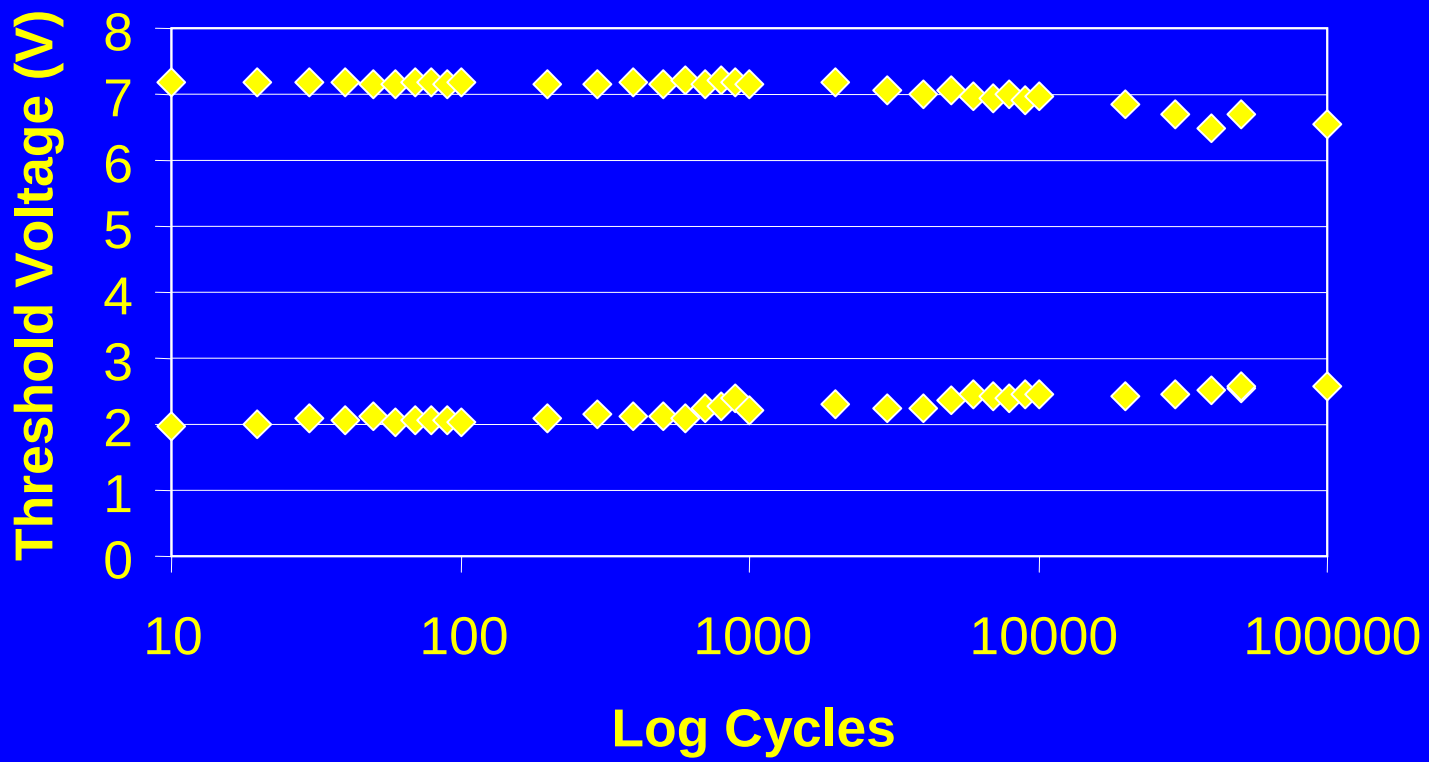
Flash Cell Charge Retention

- 500 hr bake retention data at 140 C after 10k program/erase cycles



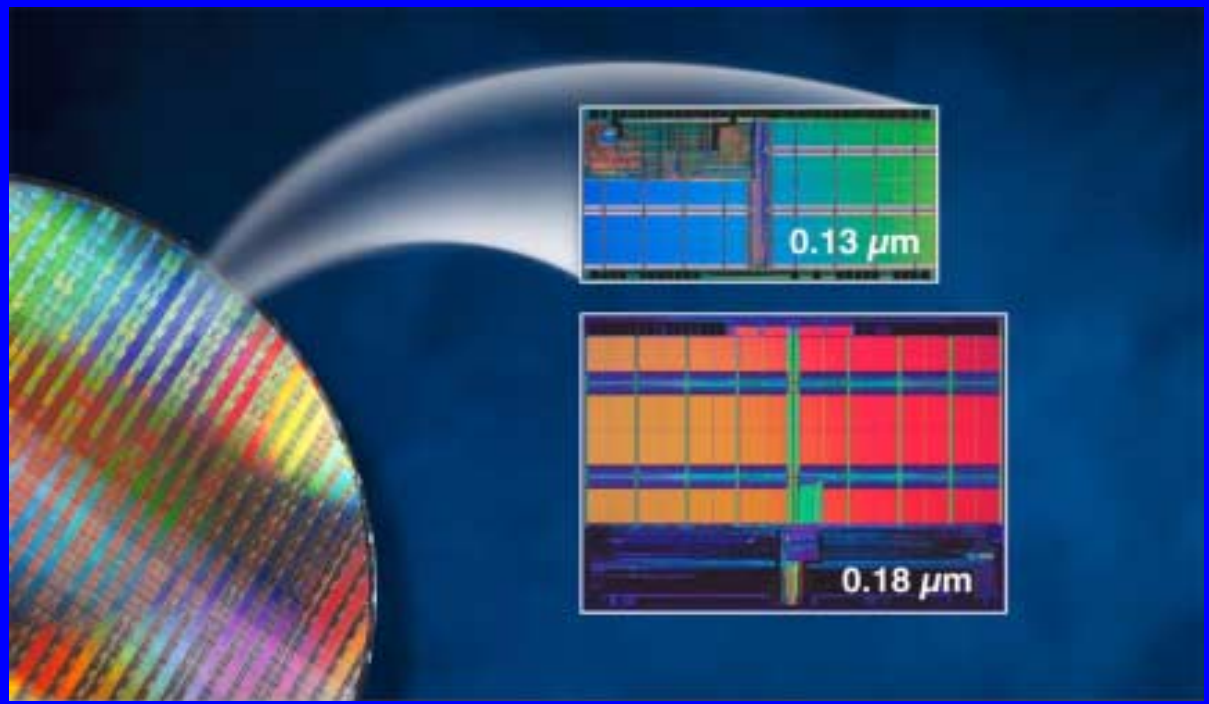
Cell Endurance Data

- **Program and erase cycling for a typical bit shows good endurance beyond 100k cycles**



Die Size Comparison

- Below is a comparison of Intel's 180nm 32Mbit product compared to the 130nm 32Mbit product showing a die size reduction of ~50%.



Note: This is a 1bit/cell product

Outline of Presentation

- Introduction
- Scaling Challenges
- Flash cell scaling
- Transistor Optimization
- MLC capability and embedded logic
- Results
- **Conclusions**
- Acknowledgements

Conclusions

- The scaling challenges for the 130nm technology were outlined.
- The scaling techniques were described resulting in a 0.16um^2 flash cell area.
- The technology was shown to be compatible with a standard logic process
- Good performance and reliability was demonstrated.
- 32Mbit product die size was shown to be reduced by 50% from the previous generation.

Outline of Presentation

- Introduction
- Scaling Challenges
- Flash cell scaling
- Transistor Optimization
- MLC capability and embedded logic
- Results
- Conclusions
- Acknowledgements

Acknowledgements

- The author would like to thank California Technology and Manufacturing (CTM) as well as the Flash Products Group (FPG) for their contribution to this development program